

# Deutsch-französische Verständigung

## Ein Übersetzungswerkzeug für das 21. Jahrhundert

von Wolfgang Teubert

Europa rückt zusammen. Mehr und mehr Menschen stehen privat und beruflich, mündlich wie schriftlich im Kontakt mit Menschen, die eine andere Sprache sprechen. Für eine kleine polyglotte Elite war das immer schon Alltag. Wer mehrsprachig aufwächst, ist im Vorteil, ebenso die, die eine natürliche Begabung für Fremdsprachen mitbringen. Meinesgleichen dagegen flößt Multilingualität erst einmal Angst ein. Spricht mich jemand in einer der vielen kleinen und auch großen europäischen Sprachen an, die mir fast alle ziemlich fremd geblieben sind, gerate ich oft in eine Panik, die schon das Erkennen abblockt, bevor es erst um das Verstehen geht. So traurig es ist, die Mehrzahl der Europäer sind weithin einsprachig geblieben, mit mehr oder weniger oberflächlichen Grundkenntnissen in einer oder selten zwei Fremdsprachen. Viele verzichten deshalb auf die Freiheit, unorganisiert ins Ausland zu fahren oder in ausländischen Sendern (oder heute im Internet) Wissenwertes zu erfahren, was der einheimische Medienmarkt vorenthält.

Erst seit dem 18. und 19. Jahrhundert halten wir Einsprachigkeit für das Normale. Damals wurden die Landessprachen als Amtssprachen durch radikale Vereinheitlichung aus den Mundarten geschaffen und die zuvor machtvollen und eine großflächige Verständigung behindernden Dialekte als Folklore gebändigt. Nichtseßhaftes Volk, Söldnerhaufen, Händler, Handwerker, Menschen auf der Flucht durchzogen die Länder und waren allgegenwärtig. Sie sprachen fremde Sprachen und unverständliche Dialekte. Die Menschen waren damals darauf eingestellt. Heute gilt: Je größer das Land, aus dem einer kommt, desto geringer seine Fremdsprachenkenntnisse. Wir halten es für die Normalität, daß jeder, mit dem wir privat oder beruflich zu tun haben, unsere Sprache spricht. Wohin wir schauen - überall spricht man deutsch in diesem großen Land. Angehörigen kleinerer Nationen scheint indessen das Fremdsprachenlernen leichter zu fallen, wie auch den Mitgliedern der unzähligen sprachlichen Minderheiten in Europa. In Luxemburg spricht jeder, so hat es wenigstens auf den ersten Blick den Anschein, ohne Mühe vier Sprachen, zählt man das Letzeburgische mit. Die Wirklichkeit ist nicht ganz so rosig. Es handelt sich eher um unterschiedliche Lebensbereiche. Französisch ist die Sprache der urbanen Gesellschaft, Letzeburgisch spricht man in Familie und der heimatlichen Gemeinschaft, die Zeitung liest man auf Deutsch, und Englisch spricht man mit den Fremden. Die Sprachen beherrscht man, wie man sie eben braucht - wer mehr will, muß sich das erhebliche Anstrengungen kosten lassen.

Für die Mehrheit der Fälle jedoch gilt, daß sich die Grenzen der Sprachbeherrschung nur mit besonderen Hilfsmitteln überwinden lassen. Notwendigkeit und Übung sind zwar erfolgreiche Lehrmeister. Man kann dies bei den Beamten der EG-Kommission beobachten, die in kurzer Frist lernen, den auftretenden Grunderfordernissen in der jeweils benötigten Sprache gerecht zu werden. Den professionellen Umgang mit multilingualen Situationen kann man lernen. Doch stößt die Sprachfertigkeit, die man dabei erwirbt, immer wieder an ihre Grenzen. Die Technik der Bewältigung multilingualer Anforderungen ist auf Werkzeuge angewiesen, auf Wörterbücher und auf Übersetzungsdienste. Alle wichtigen Verlautbarungen und vor allem die regulativen Texte der Europäischen Kommission sind durch professionelle Übersetzer in die Amts-

sprachen der Union übertragen worden - ein zwar aufwendiges, aber durch nichts zu ersetzendes Verfahren. Doch handelt es sich da nur um einen Bruchteil der gewaltigen Textproduktion europäischer Organisationen und Behörden. Übersetzungsdienste verursachen immense Kosten; deshalb reicht ihre Kapazität bei weitem nicht aus, auch die Myriaden von Gebrauchstexten in jeder ihrer zahllosen Bearbeitungsstufen zu betreuen. Enorme Verzögerungen wären sonst an der Tagesordnung. Sie stünden in eklatantem Widerspruch zu dem Zeitdruck, unter dem die Texte, die ja meist mehrere Autoren unterschiedlicher Muttersprache haben, kompiliert werden müssen. In den allermeisten Fällen gibt es für die Referenten deshalb keine Alternative dazu, selber den Text in einer Fremdsprache, meist Englisch oder Französisch, zu schreiben. Das Wörterbuch ist immer noch, wie gesagt, ihr wichtigstes Hilfsmittel.

## Werkzeuge für die multilinguale Textproduktion

Doch haben wir schon in der Schule gelernt, zweisprachigen Wörterbüchern zu mißtrauen. Das ist so nicht berechtigt. Unsere größten Wörterbücher reichen meist schon aus, um aus der Fremd- in die Muttersprache zu übersetzen. Will man sie aber benutzen, um Texte in einer Sprache zu verfassen, die man hinreichend, aber nicht perfekt gelernt hat, stößt man stets an ihre Grenzen. Zum einen bieten sie Optionen an, zwischen denen man nicht begründet auswählen kann; andererseits lassen sie uns im Stich, wenn wir etwa die Entsprechungen für Präpositionen in den allgegenwärtigen deutschen Präpositionalattributen suchen. Gerade weil der Vielfalt deutscher Präpositionen im Französischen eigentlich nur *à* und *de* gegenüberstehen, ist der Übersetzer auf Hilfe angewiesen, eine Hilfe, die ihm jedoch das Wörterbuch (auch das einsprachige übrigens) oft verweigert. Dahinter steckt weder Absicht noch Unfähigkeit der Lexikographen, sondern vor allem schlichte Raumnot. Wollte man die jedem Substantiv spezifische Syntax der Attributbildung ebenso wie die spezifischen syntaktischen Eigenschaften anderer Wörter samt ihren Übersetzungsäquivalenzen auch nur einigermaßen umfassend verzeichnen, müßte unser Wörterbuch für den Allgemeinwortschatz von 20 000 bis 30 000 Wörtern einen Umfang von leicht 10 000 Seiten haben, wie wir aus neueren Untersuchungen recht genau wissen. Das wäre ein unhandliches, vielbändiges Monstrum, dessen Benutzung höchst zeitaufwendig wäre. Kaum einer hätte das Geld, es zu kaufen.

Aber das herkömmliche Wörterbuch ist nicht nur unvollständig, sondern auch unzuverlässig. Es beruht auf den Beobachtungen und Erkenntnissen der Lexikographen und ihres Hilfspersonals. Empirische Grundlage sind in erster Linie andere Wörterbücher, daneben dann die Zettelkästen, in denen alles, was den Mitarbeitern auffällig erschien, verzeichnet ist. So heißt es beispielsweise im Vorwort von Langenscheidts Taschenwörterbuch Englisch-Deutsch von 1929: »Die Grundlage der Neubearbeitung bildet das große *Oxford English Dictionary*, ergänzt durch das *Pocket Oxford Dictionary* und eine umfangreiche eigene Sammlung von Lese- fruchten des Verfassers«. Selbst in der so fortschrittlichen angel-



sächsischen Lexikographie hat sich daran nur wenig geändert. Auch heute noch lassen dort selbst renommierte Verlage pensionierte Lehrer Neologismen suchen. Mögen Wörterbücher also gern viel Auffälliges verzeichnen, der wirklichen Wortverwendung, wie wir sie in typischen Texten der Allgemeinsprache finden, schenken sie nicht genug Aufmerksamkeit. Der Benutzer ist unzufrieden. Von der neuen Linguistik hätte er sich nützlichere Werkzeuge versprochen. Wo bleibt der Einsatz moderner Technologien? Nun gibt es im Computerzeitalter Alternativen zum herkömmlichen Buch, und gerade auch Alternativen zum Wörterbuch, das in gedruckter Form durch die Linearität des Mediums (von Ausnahmen abgesehen) auf das Alphabet als Strukturprinzip angewiesen ist. Das Einzelwort ist die einzige Zugriffsmöglichkeit, obwohl doch jeder weiß, daß sich kein Text Wort für Wort übersetzen läßt. Der Computer läßt indes einen flexiblen, interaktiv gestaltbaren Zugriff auf das gespeicherte Wissen zu, und er stellt dieses Wissen auf Knopfdruck am Bildschirm bereit. Wer einmal mit dem Oxford English Dictionary oder mit dem Grand Larousse auf CD-Rom gearbeitet hat, ist an einer vielbändigen gedruckten Version kaum noch interessiert (von Stromausfall einmal abgesehen).

Neue Werkzeuge, die die Produktion von Texten in Fremdsprachen auch dem Menschen mit durchschnittlichen Sprachkenntnissen ermöglichen, sind heute nicht nur vorstellbar, sondern machbar. Sie werden nicht billig sein. Aber das Geld für sie ist besser angelegt als jede Mark für die automatische Übersetzung, die, wo immer es um Allgemeinsprache und nicht um hochgradig formalisierte und standardisierte Fachsprache geht, eine Utopie bleiben wird. Denn ebensowenig wie ein Mensch, ausgestattet mit allen sprachwissenschaftlichen Hilfsmitteln, Texte richtig übersetzen kann, die er nicht versteht, kann das ein Computerprogramm. Zum Verständnis gehört jedoch, sich den sprachlich vermittelten Inhalt anschaulich zu machen, eine kognitive Fähigkeit, die der Computer nicht einmal simulieren kann. Deshalb wird Übersetzen, von Routinejobs abgesehen, immer eine menschliche Tätigkeit bleiben. Und deshalb werden Werkzeuge, die das Übersetzen und das Schreiben in einer fremden Sprache leichter machen, dringend benötigt. Mögen auch literarische Übersetzungen stagnieren, so führt andererseits die Globalisierung des Medienmarktes hier zu einem exponentiellen Wachstum des Übersetzungsmarktes. Meldungen von Nachrichtenagenturen, Beiträge in Publikumszeitschriften, Sach- und Fachbücher, Seifenopern und Krimiserien, touristische Gebrauchsliteratur, Bedienungsanleitungen und Enzyklopen, alles wird umgehend übersetzt oder gleich in der Fremdsprache geschrieben.

Multilingualität, wie wir sie in Europa vorfinden, und wie wir sie uns als wertvolles Kulturgut auch bewahren sollten, braucht kein Fluch zu sein. Wir sollten sie als Herausforderung annehmen und die Hilfsmittel schaffen, mit denen sich multilinguale Situationen bewältigen lassen. Denn die technischen Voraussetzungen sind inzwischen geschaffen. Jetzt bedarf es der großen Anstrengung, den Schritt ins 21. Jahrhundert zu wagen.

## Warum Deutsch-Französisch, Französisch-Deutsch?

Unser neuartiges Übersetzungswerkzeug wird viele Millionen Mark kosten. Das ist weit mehr Geld, als Verlage für zweisprachige Wörterbücher normalerweise ausgeben. Andererseits ist es nur ein Bruchteil dessen, was die EG-Kommission und die Mitgliedstaaten in den achtziger Jahren in das völlig gescheiterte EUROTRA-Projekt mit dem Ziel einer automatischen Übersetzung investiert haben und was heute die Bundesregierung für das Großvorhaben VERBMÖBIL mit dem noch ehrgeizigeren Ziel einer vollautomatischen Übersetzung gesprochener Sprache ausgibt.

Auf den ersten Blick drängt sich das Sprachenpaar Englisch-Deutsch für ein solches technisch innovatives Vorhaben auf. Ist

nicht Englisch die Weltsprache schlechthin, ist sie nicht längst zur lingua franca, zur Interlingua für die geworden, die sich nicht direkt mit dem Nachbarn sprachlich verständigen können? Muß nicht, wer international zur Kenntnis genommen werden möchte, sich ohnehin des Englischen bedienen? Können oder sollen wir nicht die direkte Verständigung mit unseren anderen europäischen Nachbarn der kleinen polyglotten Elite überlassen, in deren Händen es ja immer schon ruhte? Jedenfalls würde das teure Übersetzungswerkzeug für Deutsch und Englisch einen radikal größeren Markt finden als für Deutsch und Französisch. Warum dann nicht gleich Deutsch-Finnisch?

Die europäische Multilingualität als kulturellen Wert zu bewahren kann nicht bedeuten, daß wir künftig von jedem Bürger Europas erwarten, sich in allen Partnersprachen verständigen zu können. Das ist eine Utopie, und das kann nicht die Alternative zur völligen Unterwerfung unter eine einzige Weltsprache sein. Nun ist die Existenz von Weltsprachen nicht neu. Im Hellenismus gab es die Koine, im europäischen Mittelalter das Latein; seit der Aufklärung Französisch und Englisch. Das waren die Interlinguen in der kleinen eurozentrischen Welt damals, derer man sich bediente, wenn man nicht direkt miteinander ins Gespräch kommen konnte.

Aber neben diesen Weltsprachen gab es immer auch überregionale Verkehrssprachen. In Europa gehörten dazu, bei unterschiedlichen Voraussetzungen und zeitlich wie situativ manchen Einschränkungen unterliegend, Spanisch, Russisch, Italienisch und Deutsch. Heute müssen wir dazu allmählich auch Französisch rechnen, was die kontinentaleuropäischen Verhältnisse betrifft. Auch wenn Französisch weiterhin global gesehen Weltsprache bleiben wird, ist es in Europa nur eine unter mehreren überregionalen Verkehrssprachen, sicher in einer herausgehobenen Rolle. Diese Rolle verdankt sich seiner Funktion als klassische europäische Bildungssprache. Es soll hier nicht diskutiert werden, wieweit Spanisch in Europa noch als überregionale Verkehrssprache gelten kann. Italienisch gehört sicher noch dazu: Österreicher, vielleicht auch Bayern, Südslawen und wohl auch ein Teil der Griechen ziehen es dem Französischen vor. Ob das während des Sowjetimperiums so unbeliebt gewordene Russisch wieder einmal zur Verkehrssprache unter den slawischen Völkern wird, bleibt abzuwarten.

Überregionale Verkehrssprachen sind die Sprachen zahlenmäßig großer Länder; wer eine oder zwei überregionale Verkehrssprachen beherrscht, kann sich in einer statistisch sehr erheblichen Zahl bilateraler Kontakte direkt mit seinem fremdsprachigen Partner verständigen.

Ließe sich der Anspruch europäischer Vielsprachigkeit dadurch befördern, daß man den überregionalen Verkehrssprachen eine besondere Rolle zuweist? Würden damit nicht die Sprachen kleinerer Nationen in eine dritte Klasse herabgewürdigt? Letzteres ist vordergründig gewiß richtig. Aber es wäre unrealistisch zu erwarten, daß es jemals eine Massenbewegung von Deutschen, die Dänisch, oder von Italienern, die Slowenisch lernen, geben könnte. Slowenen gar, die Dänisch, und Dänen, die Litauisch lernen, gehören unbedingt zu einer kleinen polyglotten Elite; um ihren Bestand war und ist nicht zu fürchten. Worum es geht, ist eher die Bestärkung der Vorstellung, daß wir aus der kulturellen Vielfalt Europas nur Gewinn ziehen können, wenn wir die kulturellen Unterschiede möglichst direkt, ohne die Vermittlung über eine dritte Kultur oder eine dritte Sprache, wirken lassen. Was an einem italienischen Text so typisch italienisch ist, läßt sich eben nur in der direkten, bilateralen Verständigung gut nachvollziehen. Kulturelle Vielfalt bedeutet ein Angebot von Optionen, Dinge auch anders zu sehen und anders zu machen, als man das von zu Hause gewohnt ist; sie bietet den Ausweg aus der allgegenwärtigen Tendenz, die überkommene Organisation der Gesellschaft, das Bildungs-, Arbeits- und Sozialsystem einzubetonieren, ohne die geänderten Rahmenbedingungen zur Kenntnis zu nehmen. Kulturelle Vielfalt begünstigt Innovation, und ein fremdes Beispiel wird leichter zum Vorbild, wenn wir



direkt mit den Akteuren in ihrer oder unserer, nicht aber in einer dritten Sprache sprechen.

Die besondere gegenseitige Förderung überregionaler Verkehrssprachen hat deshalb nichts mit hegemonialem Denken zu tun. Sie ist vielmehr die Voraussetzung dafür, daß möglichst viele Verständigungsakte in Europa direkt, ohne Vermittlung einer dritten Sprache, stattfinden. Wer mit einem Europa, in dem es nur eine supranationale Amts- oder Arbeitssprache gibt, nämlich Englisch, nicht einverstanden ist, muß zunächst einmal den Gebrauch überregionaler Verkehrssprachen begünstigen.

Unter den Verkehrssprachen ist das Französische die Nachbarsprache des Deutschen. Wollen die Deutschen die Verständigung mit ihrem großen westlichen Nachbarn nicht nur den polyglotten Eliten überlassen und auch nicht Englisch als Interlingua akzeptieren, bleibt nur, die Bedingungen der deutsch-französischen sprachlichen Verständigung zu verbessern. Wenn die Deutschen die Franzosen und die Franzosen die Deutschen auf etwas aufmerksam machen wollen, sei es, um dem anderen etwas zu verkaufen, sei es, um ihm Ideen, Kunst, Ordnungen oder Anschauungen nahezubringen, müssen sie sie in ihrer Sprache ansprechen. Deshalb sollte das innovative Werkzeug, das das Übersetzen und Schreiben in fremden Sprachen erleichtern soll, nicht mit dem Sprachenpaar Deutsch und Englisch, sondern mit den vergleichbaren europäischen Verkehrssprachen Französisch und Deutsch beginnen. Bewährt es sich, steht es als Modell für beliebige weitere Sprachenpaare zur Verfügung. Daß genug Bedarf besteht, auch für die sogenannten kleineren Sprachen, läßt sich in letzter Zeit verstärkt registrieren.

## Das Prinzip des Übersetzungswerkzeugs

Wie sieht es nun aus, das geplante Hilfsmittel zum Übersetzen und Schreiben in fremden Sprachen? Drei Leistungen vor allem erwarten wir:

- Es soll Übersetzungsvorschläge nicht nur für das Einzelwort, sondern für ganze Phrasen (z.B.: *einer Verfügung Folge leisten*) und für Kollokationen (z.B.: *gerichtliche Verfügung, unbedingte Folge*) anbieten.
- Es soll, was den Allgemeinwortschatz von 20.000 bis 30.000 Wörtern angeht, das syntagmatische und paradigmatische Verhalten der lexikalischen Einheiten möglichst vollständig in Hinsicht auf die Benutzerbedürfnisse erschließen. (Für den weiteren Wortschatz ist diese Erschließung nicht erforderlich, denn diese selteneren Wörter sind meist semantisch eindeutig und haben eine regelmäßige Syntax.)
- Es soll das, was bereits übersetzt ist, abspeichern und wieder anbieten, wenn die Übersetzungseinheit in gleicher oder vergleichbarer Form wiederkehrt.

Beginnen wir mit der Vollständigkeit. Anders als für gedruckte Wörterbücher ist Vollständigkeit für das elektronische Wörterbuch weder ein (Speicher-)Platzproblem noch ein Problem für die Zugriffsgeschwindigkeit. Es ist lediglich noch eine Frage verfügbaren lexikalischen Wissens. Der Zwang zur Beschränkung hat Lexikographen bisher daran gehindert, das syntagmatische und paradigmatische, das syntaktische und semantische Verhalten der Wörter vollständig zu ermitteln und darzustellen. Wörterbuchverlage können es sich kaum leisten, empirische Untersuchungen großer Textmengen zu finanzieren. Um die Kosten im Griff zu halten, müssen sie sich vielfach auf vorhandene Wissensrepositorien verlassen, auf die Kompetenz der Lexikographen und bereits existierende Wörterbücher also.

Seit einigen Jahrzehnten ist an die Stelle der lexikographischen Zettelkästen von sprachlichen Auffälligkeiten das Korpus, d.h. eine wie auch immer definierte elektronisch verfügbare Sammlung von Texten getreten. Ist es einigermaßen ausgewogen zusammengestellt, so bilden diese Texte den Allgemeinwortschatz ab. Inzwi-

schen gibt es für das Französische und für das Englische, aber noch nicht für das Deutsche, Korpuswörterbücher. Sie beschreiben den Wortgebrauch auf der Grundlage aller Wortvorkommen im Korpus, also systematisch auf Vorkommenshäufigkeit (und nicht nach Auffälligkeit) gerichtet. Korpora müssen auch Grundlage für die zweisprachige Lexikographie sein. Nirgendwo sonst als im Korpus manifestiert sich das erforderliche lexikalische Wissen.

Aber wir benötigen mehr. Für jeden Wortgebrauch in der einen Sprache suchen wir das ideale Übersetzungsäquivalent in der anderen Sprache. Wo uns die vorhandenen Wörterbücher im Stich lassen, müssen wir uns wieder an das Korpus halten, genauer, an solche Korpora, die Originaltexte in einer und deren Übersetzung in der anderen Sprache enthalten. Man nennt sie parallele Korpora, und damit man mit ihnen arbeiten kann, müssen sie erst einmal parallelisiert, d.h. Satz für Satz aufeinander bezogen werden. Das ist keineswegs trivial; in der Übersetzung werden oft aus einem Satz mehrere, oder mehrere Sätze werden zu einem zusammengezogen. Doch gibt es inzwischen bewährte Verfahren zur automatischen Parallelisierung.

Unsere Erfahrungen lehren uns, daß wir, um den Wortschatz der Allgemeinsprache zu beschreiben, ein Korpus von 100 Millionen Wörtern benötigen. In dieser Größenordnung bewegt sich das Korpus des *Trésor de la langue française* in Nancy, der *Bank of English* in Birmingham und des gerade fertiggestellten regierungsfinanzierten *British National Corpus*. Solche Korpora sind teuer, vor allem in Hinblick auf die elektronische Speicherung, und zwar auch, wenn mit Lesemaschinen eingelesen oder wenn bereits vorhandene elektronische Datenträger, etwa Setzbänder, benutzt werden. Am teuersten ist es jedoch, Texte abzuschreiben.

Ein paralleles Korpus Deutsch-Französisch müßte einen Umfang von 400 Millionen Wörtern haben, wenn es so leistungsfähig sein soll wie ein einsprachiges Korpus von 100 Millionen Wörtern Länge. Denn es besteht aus deutschen Originaltexten mit französischen Übersetzungen und aus französischen Originaltexten mit deutschen Übersetzungen. Ein solches Korpus, das zudem noch ausgewogen wäre, ist vorerst wohl unbezahlbar. Realistischer wäre ein Parallelkorpus von insgesamt 100 Millionen Wörtern, also je 25 Millionen Wörter Original und Übersetzungen pro Sprache. Dieses könnte zur Kontrolle und Ergänzung mit einsprachigen Referenzkorpora (mit einheitlichem Aufbau, wie sie gerade im Rahmen einer EU-Initiative erstellt werden) zusammengekoppelt werden. Das wäre eine erträgliche Notlösung.

Ein solches Korpus enthält weit mehr als die Summe alles in Wörterbüchern verzeichneten lexikographischen Wissens. Denn es bildet das ab, was Übersetzer im Kopf haben. Zwar kommen Übersetzer ohne Wörterbücher nicht aus. Wichtiger aber ist für sie ihr in der Übersetzungsarbeit erworbener Erfahrungsschatz, implizites Wissen, das wie die Kunst des Treppensteigens sich eigentlich der Beschreibung entzieht. Das Übersetzungswissen, das unser geplantes Werkzeug dem Benutzer anbieten soll, manifestiert sich explizit in diesen Übersetzungen. Fast alles, was es zu übersetzen gilt, wurde so oder ähnlich schon einmal übersetzt. Darunter sind gute und schlechte Übersetzungen, sogar falsche. Durch Gegenproben ist es möglich, automatisch die Spreu vom Weizen zu trennen. Inzwischen sind Verfahren in der Entwicklung, die automatisch die passenden Übersetzungsäquivalente in den Korpora ausfindig machen können: manchmal Einzelwörter, oft Kollokationen und ganze Phrasen.

In diesem aufzubauenden Parallelkorpus Französisch-Deutsch findet sich ein Schatz von Übersetzungsäquivalenten, mit dem kein traditionelles Wörterbuch mithalten kann. Im Prinzip wird man darin für 95% bis 98% aller Fälle entweder die fertige Phrase oder ein geeignetes Strukturmuster finden, das man im Analogieschluß verwenden kann. Aber wir erwarten mehr von unserem Hilfsmittel. Der Benutzer soll nicht gezwungen sein, fünfzig oder hundert oder gar fünfhundert Belege am Bildschirm auf ihre Eignung zu über-



prüfen. Wir erwarten, daß unser Werkzeug selbst die eine Lösung (oder verschiedene alternative Lösungen) anbietet, die als Übersetzungsäquivalent konkret in Frage kommt. Für einen Satz, der aus dem Deutschen ins Französische zu übersetzen wäre, muß, bezogen auf ein Wort dieses Satzes, der Beleg im Parallelkorpus ermittelt werden, der dem Satz möglichst weitgehend entspricht, der in Hinblick auf die lexikalische Füllung der Satzglieder möglichst weitgehend übereinstimmt.

Ein Beispiel wäre der zu übersetzende Satz mit dem Wort *Bericht* als primärem Suchbegriff: *Die Agentin erstattete dem Oberst Bericht von ihrer Begegnung mit dem Anführer*. Der Korpusbeleg, der am nächsten kommt, sei: *Er erstattet dem Präsidenten Bericht von den Verhandlungen mit der Gewerkschaft*. Er ist übersetzt als: *Il rend compte au Président des pourparlers avec le syndicat*. Wird er durch die *Agentin*, von den *Verhandlungen mit der Gewerkschaft* durch von *ihrer Begegnung mit dem Anführer, dem Präsidenten* durch dem *Oberst* ersetzt und das Prädikat ins Präteritum gesetzt, erhält der Benutzer für den ganzen Satz komplett das Übersetzungsäquivalent: *l'agent a rendu compte au colonel de son rendez-vous avec le chef*. Alternativ könnte zu *rendez-vous* auch *rencontre* angeboten werden. Davon später. Der Benutzer vermeidet so das hier unidiomatische *a fait son rapport* für *erstattete Bericht* und braucht sich keine Gedanken über die zu verwendenden Präpositionen zu machen.

Wie funktioniert das? Ein Programm analysiert die zu übersetzenden Sätze und die Korpusbelege, indem jedem Wort Wortart und die morphologischen Funktionen wie Kasus, Genus, Numerus, Tempus zugeordnet werden. Ein weiteres benutzt diese Daten, um Satzglieder, Objekte, Angaben und das Prädikat zu identifizieren. Wie eine Abfrage im gespeicherten Hintergrundlexikon ergibt, hat *Bericht* *erstatten* drei Ergänzungen: das Subjekt (im Nominativ), ein Dativobjekt und ein mit *von* oder *über* eingeleitetes Präpositionalobjekt. Ein solcher Satz soll übersetzt werden, und ein solcher Satz wird im Korpus mit Übersetzungsäquivalent gefunden. Auf ähnliche Weise werden auch die Äquivalente für *Agentin*, *Oberst* und *Anführer* gefunden. Das hier kurz angedeutete Verfahren beruht auf der Anwendung linguistischen Wissens.

Aber bei vielen Problemen fehlt uns das linguistische Wissen, das erforderlich wäre, um das passende Übersetzungsäquivalent auszuwählen. Das zeigt uns ein Blick in *Langenscheidts Großwörterbuch Deutsch-Französisch* bzw. *Französisch-Deutsch* (Sachs-Villatte, 1979), wenn wir etwa die Übersetzungen für *Kummer* und *Gram* suchen. Für *Kummer* finden wir (ohne weitere Erläuterungen): *chagrin*; *souci(s)*; plusfort. *peine*; *affliction*. Für *Gram* finden wir: *chagrin*; *peine*; *affliction*. In umgekehrter Richtung finden wir für *chagrin*: *Kummer*, *Gram*, *Betrübnis*, *Leid*; für *souci*: *Sorge*, *Besorgnis*, *Kummer*; für *peine*: *Kummer*, *Schmerz*, *Leid*, *Sorgen*; schließlich für *affliction*: *Kummer*, *Leid*, *Betrübnis*, *Bekümmernis*, *Trübsal*, *Gram*. An den fehlenden semantischen Erläuterungen zeigt sich (neben der Raumnot) die Hilflosigkeit der Lexikographen. Die gegebenen Informationen reichen zwar aus, um aus der Fremdsprache in die Muttersprache zu übersetzen, aber nicht in umgekehrter Richtung. Denn mögen die Wörter auch ähnlich in ihren Bedeutungen sein, es gibt kaum einen Kontext, in dem sie substituierbar sind. Es hängt ganz von der Situation ab, ob man das eine oder andere Wort gebraucht. Jede Sprache, jede Kultur benutzt ihre eigenen Parameter, um Bedeutungsfelder dieser Art zu differenzieren. Was *chagrin*, *peine* und *affliction* voneinander unterscheidet, hat wahrscheinlich nichts mit dem zu tun, was *Kummer* von *Gram* trennt.

Ein auf der Anwendung morphologischer und syntaktischer Wissens beruhendes Verfahren, wie wir es oben für *Bericht* *erstatten* vorgeführt haben, würde uns hier nicht weiterhelfen. In unserem Parallelkorpus finden wir sicher zahllose Belege für die genannten Wörter mit ihren Entsprechungen in der jeweils anderen Sprache. Wenn *chagrin* danach zu 70% mit *Kummer* und zu 30% mit *Gram* übersetzt wird, welche Alternative soll das elektronische Wörter-

buch dann anbieten? Eine Möglichkeit wäre der Zugriff auf vorhandene einsprachige Wörterbücher, die man elektronisch verfügbar machen und dem geplanten Werkzeug zuschalten könnte. Davon soll abgesehen werden, zum einen, weil solche semantischen Differenzierungen nirgends systematisch durchgeführt sind, zum anderen, weil man sich schlechterdings nicht auf sie verlassen kann. Denn die vorhandenen Wörterbücher sind nicht korpusbasiert.

Wir müssen also ein Verfahren entwickeln, das den Bedeutungsunterschied (oder sollte man besser sagen: Verwendungsunterschied) von *Kummer* und *Gram* im Korpus (diesmal im deutschen Korpus) ausfindig macht. Dieses Verfahren würde etwa untersuchen, wie sich in den Belegen die jeweiligen Kontexte (z.B. zehn Wörter links und rechts vom Suchwort) in statistischer Analyse unterscheiden. Gibt es bestimmte Wörter, die in signifikanter Häufigkeit im Kontext von *Gram* vorkommen? Bei *Kummer* werden wir vermutlich besonders häufig *jung*, *Mädchen*, *sie*, *ihr* usw., bei *Gram* eher *alt*, *Mann*, *er*, *sein* finden. Ein solches Verfahren läßt sich vollautomatisch durchführen. Mit diesem Befund ist nun der zu übersetzende französische Satz abzugleichen, beispielsweise: a) *Dans son chagrin la jeune fille se retirait de ses amies*. Die Entsprechungen von *elle* und *amies* sind *sie* und *Freundinnen*, letzterer ein Unterbegriff des Begriffs 'weibliche Person', der auch in *Mädchen* steckt. Angesichts des Fehlens gegenteiliger Evidenz kann unser elektronisches Wörterbuch beruhigt *Kummer* als Übersetzungsäquivalent von *chagrin* anbieten. Anders in dem Satz: b) *Dans son chagrin la jeune fille se retirait de son vieux père*. Hier gibt es im Rahmen des oben geschilderten Verfahrens ähnlich starke Indizien für beide Alternativen. Einerseits finden wir *sie*, *ihr*, aber andererseits finden wir auch *alt* und in *Vater* einen Unterbegriff zu 'männlicher Person'. Hier muß das System den Benutzer fragen, welche der beiden Verwendungsmuster (jung/weiblich bzw. alt/männlich) im Kontext von *chagrin* relevant gesetzt werden sollen.

Statistikbasierte Verfahren wie das hier vorgestellte gibt es bereits. Obwohl sie ohne linguistisches (syntaktisches oder semantisches) Wissen auskommen, haben sie eine erstaunlich hohe Trefferrate. Sie können weiter optimiert werden, wenn sie um eine syntaktische Komponente ergänzt werden. So ist es in den meisten Belegen das Subjekt im Nominativ, das *Kummer* *hat*, oder es *empfindet* *Gram*, und in den meisten anderen Belegen finden wir ein Possessivpronomen unmittelbar vor dem Suchwort oder ein Genitivattribut unmittelbar danach: *ihr Kummer* oder *der tiefe Gram des alten Vaters*; und statistisch relevant kommt es nur auf diese beiden Positionen Subjekt und Nominalattribut an.

Damit ist angedeutet, wie unser geplantes Übersetzungswerkzeug die ersten beiden der oben genannten Leistungen erbringt: einerseits Übersetzungsvorschläge für ganze Phrasen, nach Möglichkeit ganze Sätze, zum anderen die Erschließung von Bedeutungsunterschieden zwischen Übersetzungsalternativen. Zusätzlich verlangen wir von dem Werkzeug aber auch die Fähigkeit, sich an das, was bereits übersetzt (oder, falls direkt in der Fremdsprache geschrieben: vom System abgesegnet) ist, zu erinnern. Diese Komponente nennen wir Übersetzungsgedächtnis, *translation memory* im Jargon der Experten. Eine ganze Reihe von Prototypen wurden bereits erfolgreich entwickelt. Translation memories reduzieren den Übersetzungsaufwand um bis zu 50% und sind damit anderen teilautomatischen Übersetzungsverfahren im Preis-Leistungs-Verhältnis konkurrenzlos überlegen. Sie beruhen auf dem simplen Gedanken, daß in einem Text von beispielsweise 500 Seiten Länge 90% aller Wörter und Phrasen auf den ersten 50 bis 100 Seiten vorkommen und danach nur in stets neuen Kombinationen und Varianten wiederholt werden. Mit diesem Werkzeug wird aber die Übersetzungsarbeit nicht nur schneller, sondern mit ihr kann man auch die Ambiguitäten und Vieldeutigkeiten des Originaltextes reduzieren. Wenn das oben beschriebene Statistikverfahren bereits in fünf Fällen *chagrin* als *Kummer* wiedergegeben hat, ist diese Übersetzung aller Wahrscheinlichkeit nach auch im sechsten Fall richtig, wo der Kontext keine Festlegung zuläßt - es



sei denn, in der Folge finden sich Fälle, wo *chagrin* mit großer Sicherheit als *Gram* wiederzugeben wäre. Auch in der oben erörterten Übersetzung von *Bericht erstatten* hatten wir es mit einer Ambiguität zu tun: *Treffen* können zufällig oder verabredet sein, und dafür gibt es im Französischen zwei Wörter: *rencontre* bzw. *rendezvous*. Man darf erwarten, daß in einem längeren Text irgendwo (üblicherweise beim ersten Auftreten) der Kontext die Verwendung etwa durch die Kombination mit *verabreden* vereindeutigt worden ist. In Folgefällen, in denen der Kontext neutral ist, liegt es dann nahe, bei der zunächst gewählten Alternative zu bleiben.

Die Darstellung der Arbeitsweise der vielen Verfahren und Komponenten des geplanten Übersetzungswerkzeugs hat hoffentlich deutlich gemacht, daß dieses Hilfsmittel nicht einfach ein elektronisches Wörterbuch und überhaupt keine Auflistung vorgefertigter Lösungen, sondern im Gegenteil eine komplexe Konfiguration einer Anzahl von aufeinander bezogenen Prozeduren und Teilsystemen in Verbindung mit einem sehr großen Korpus ist. Es enthält keine fertigen, auf definierte lexikalische Einheiten bezogenen Übersetzungsregeln oder Wörterbucheinträge. Unsere Lexikographen schreiben kein Wörterbuch mehr. Sie entwickeln zusammen mit Informatikern und Statistikern automatisch ablaufende Verfahren und konfigurieren sie in einer Architektur so, daß dieses elektronische Werkzeug herkömmliche Wörterbücher nicht nur ersetzt, sondern bei weitem übertrifft. Der Benutzer aktiviert sein Werkzeug, das mit seiner Arbeit beginnt, sobald der erste zu übersetzende Satz eingegeben ist. Im Zusammenspiel von statistischen und linguistisch determinierten Verfahren macht es sich im Parallelkorpus auf die Suche nach einem Vergleichsfall und überträgt ihn auf die gestellte Aufgabe. Im Idealfall bietet es phrasen- oder satzweise fertige Übersetzungsäquivalente an. Sonst konfrontiert es den Benutzer mit Alternativen, unter denen er auswählt. Die angebotenen Lösungen beruhen auf der Analyse der Kontextdaten und sind damit jedem fertigen Wörterbuchwissen überlegen.

Voraussetzung ist eine Technik, die die Ergebnisse so schnell liefert, daß der Benutzer nicht ungeduldig wird. Prinzipiell gibt es da kein Problem. Zwar wäre ein normaler PC zu klein, nicht aber eine *work station*. Trotzdem ist es ein ehrgeiziges Vorhaben. Der Ersatz des traditionellen statischen Wörterbuchs mit verallgemeinerbaren Übersetzungsregeln durch ein dynamisches System, das fallspezifische Übersetzungshilfe leistet, bedeutet einen Paradigmenwechsel in der Übersetzungsarbeit. Die bi- und multilinguale Lexikographie des 21. Jahrhunderts hat ein neues Gesicht. Sie muß es haben, um den gewaltig gestiegenen Anforderungen gerecht zu werden.

Was wir hier vorgestellt haben, ist ein überaus ehrgeiziges Konzept. Die einzelnen Komponenten, aus denen sich unser Übersetzer-Werkzeug zusammensetzt - und es sind einige mehr, als wir hier beschrieben haben - liegen zwar alle vor, wenn auch teilweise noch im Experimentalstadium. Die Kunst liegt in der Gesamtarchitektur des Systems, in der optimalen Verzahnung der Prozesse, so daß die Ergebnisse dem Benutzer möglichst ohne Verzug dargeboten werden. Das ist im wesentlichen ein Informatikproblem. Dagegen ist die sich auf der Grundlage von automatischer Kontextanalyse selbst organisierende semantische Klassifikation von lexikalischen Einheiten, für die es in der Zielsprache zwei oder mehrere Übersetzungsäquivalente gibt, eine Herausforderung an die Linguistik. Gründliche Planung tut not. Vor einer Entscheidung müssen wir uns vergewissern, ob die Zeit schon reif ist für dieses elektronische Übersetzungshilfsmittel der Zukunft, oder ob man den heutigen Anforderungen ein letztes Mal mit einem nun zwar

korpus-basierten, sonst aber traditionellen zweisprachigen Wörterbuch in elektronischer Form entsprechen sollte.

## Zum Stand der Vorbereitungen

Daß das Institut für deutsche Sprache gerade jetzt konkrete Überlegungen zu oben anstehenden Aufgaben der deutsch-französischen Lexikographie anstellt, kommt nicht von ungefähr. Schon seit längerem hat es die Kontakte zum Partnerinstitut in Frankreich, dem *Institut National de la Langue Française* (INaLF, Paris und Nancy) intensiviert. Seit einigen Jahren gibt es bereits gemeinsame Vorhaben, die den Einsatz des Computers für die zweisprachige und multilinguale Lexikographie vorbereiten. Auch in der Neologieforschung ist es inzwischen mit dem am INaLF assoziierten *Laboratoire Linguistique Informatique* (Leitung: Gaston Gross) zu einer auf Dauer angelegten Zusammenarbeit gekommen.

In einem Gespräch im Frühjahr 1995 haben Gerhard Stickel, Direktor des IDS, und Bernard Quemada, Vize-Präsident des direkt dem Ministerpräsidenten unterstellten *Conseil de la langue française* und Nestor der französischen Computerlinguistik, sich über die Notwendigkeit eines neuen, korpus-basierten deutsch-französischen elektronischen Übersetzungswerkzeugs verständigt. Mit Robert Martin, dem Direktor des INaLF, wurde eine Zusammenarbeit zwischen beiden Instituten vereinbart.

Diese Initiative traf sich mit ähnlichen Plänen von Franz Josef Hausmann, Lexikographieexperte an der Universität Erlangen, und Jean-Marie Zemb, Germanist und Spezialist für deutsch-französischen Sprachvergleich am Collège de France in Paris, die diese Idee schon seit längerem verfolgen und für die Realisierung ein Zusammengehen von INaLF und IDS vorgeschlagen haben. Inzwischen fand, am 19. und 20. Juni 1995, ein Kolloquium am Collège de France statt, in dem führende Lexikographen, Computerlinguisten, Vertreter von Wörterbuchverlagen und Organisationen, die sich in der Förderung deutsch-französischer Zusammenarbeit engagieren, ihre Erwartungen an das dringlich geforderte neue Übersetzungswerkzeug konkretisiert haben und über Konzepte und Modell diskutierten.

Das geplante Vorhaben wird nicht billig sein. Erforderlich sind sicher 80 bis 100 Arbeitsjahre für Sprachwissenschaftler und Informatiker, dazu umfangreiche Mittel für Korpusaufbau und Programmierung, nach heutigen Kosten ca. 15 Millionen Mark. Ein Projekt dieser Größenordnung setzt die Bildung eines arbeitsfähigen Konsortiums von Forschungsinstituten, Softwareindustrie und Wörterbuchverlagen in beiden Ländern voraus. Als innovative Unternehmung im Bereich Sprachtechnologie leistet dieses Vorhaben einen Beitrag dazu, die französische und deutsche Sprachindustrie auf einem wichtigen Zukunftsmarkt konkurrenzfähig zu machen. Dies sollte den öffentlichen Händen beider Länder eine großzügige Förderung wert sein. Mit der gezielten Förderung von Französisch und Deutsch als zwei benachbarten überregionalen Verkehrssprachen trägt das Projekt zum Konzept eines Europas der kulturellen und sprachlichen Vielfalt bei.

Der Autor ist Leiter der Abteilung Sprachentwicklung in der Gegenwart am Institut für deutsche Sprache (Mannheim).